网购评语高频词共现网络的结构特征分析

李桃迎，吕晓宁，李　峰

(大连海事大学 交通运输管理学院，辽宁 大连 116026)

摘　要：网购评语是消费者对网购商品的直接反馈，从中挖掘有价值的知识有助于为商家开展精准化营销和个性化推荐服务、消费者制定购买决策等提供依据。鉴于此，以国内大型综合型电商平台上服装类网购评语为研究对象，对评语分词、筛选高频词，分析高频词之间的共现关系，构建高频词共现网络，分析得出网络评语的热点词多个结构特征和评语网络中少数节点对网络的运行起着主导的作用，为网购评语挖掘研究领域提供了按照网购评语高频词共现网络的结构特性对销量的交互影响进行研究的思路。

关键词：网购评语；高频词；共现网络；情感分析

中图分类号：TP391　　doi: 10.3969/j.issn.1001-3695.2017.06.0657

Analysis of structure characteristics of high frequency word co-occurrence network of online shopping reviews

Li Taoying, Lyu Xiaoning, Li Feng

(*Transportation Management College*, *Dalian Maritime University*, *Dalian Liaoning* 116026, *China*)

Abstract: Consumers' online shopping reviews are consumers' feedback to online shopping. Mining valuable knowledge from massive online shopping reviews will not only provide safeguard for businesses to carry out precision marketing and personalized recommendation services, but also is good for consumers to make purchase decisions. Besides, management departments can use it to establish regulatory strategy. In this paper, the clothing online shopping comments on China's large integrated electronic business platform is reviewed as the object of study, making a participle of the comments, screening high frequency words, analyzing of co-occurrence relationship between the high frequency words in order to structure high frequency word co-occurrence network, a number of structural features for the network comments hot words is shown by analyzing the network and in the comment network, a few nodes play a dominant role in the operation of the network. On the basis of the conclusion, this research also provides the study suggestion on the interaction effects of the structure characteristics of the high frequency word co-occurrence network on the sales volume.

Key Words: online shopping reviews; high frequency word; co-occurrence network; sentiment analysis

0　引言

互联网的极速发展，现代信息化的迅速普及，使得网络购物成为高效快捷的购物方式，广泛影响着人们的日常生活和工作。随着网上购物人数的与日俱增，在线商品评论的数量不断增加，潜在消费者越来越难从中发现有助于制定购买决策的信息[1]。同时每个消费者在评论商品的时候，由于用语习惯的不同，会导致评论的无组织性和非结构化。并且每条评语对商品的评价涉及到不同方面，顾客想制定购买决策、电商企业想提高消费业绩、监管部门监督管理都需要从大量评语中寻找自己的关注点，这将会十分的不易。直接浏览商品的评论信息不仅耗时耗力，也很难形成一个客观整体的印象，对于顾客、电商企业、监管部门而言，是一个极其繁琐的过程。浏览评语的目的主要想了解商品某一具体属性的信息，如外观、质量、服务态度等。各大电子商务网站的评论功能，给了消费者了解商品实际情况的信息渠道和平台，其及时、便捷、互动的特性满足了顾客的信息需求。所以如何快速的从繁琐大量的信息中提取出对客户、商家、监管部门有价值的信息就显得至关重要。

网购评语挖掘相关的文献有很多，主要集中在三个方面：

a）评语的真实性与有用性[2~8]。如廖成林等人[2]运用实证研究方法，探讨了在线商品评论有用性的影响因素以及各因素之间的作用机制，姜巍等人[4]认为评论质量的良莠不齐严重干扰了需求挖掘的准确性和可信性，提出一种基于复杂网络的评论有用性分析方法[5,6]，消费者发表在线评论的意愿和参与动机

作者简介：李桃迎（1983-），女，安徽宿州人，副教授，博士，主要研究方向为数据挖掘、网购行为分析研究（litaoying@dlmu.edu.cn）；吕晓宁（1992-），女，河北衡水人，硕士研究生，主要研究方向为社会网络分析；李峰（1993-），女，山西朔州人，硕士研究生，主要研究方向为文本挖掘、情感挖掘.

等行为影响因素[7]，如小额经济回报等不同方式的激励措施，都鼓励消费者发表在线产品评论，并可能作出偏离真实性的评论[9]，甚至出现虚假评论信息，使得商品评论信息可信度受到严重影响[10]。这类文献主要探讨评论是否真实反映了消费者的真实意愿以及评论的质量，即评价的有效性。

b）评语对购买决策或销量的影响[13-23]。如王君珺等人[13]研究了热门品牌产品的在线评论对非热门品牌产品销量是否具有影响，李宏等人[14]研究了负面在线评论质量、消费者卷入度和性别三个变量对消费者满意度和购买选择的影响，龚诗阳等人[21]通过当当网图书的评论数据进行实证分析，发现线上消费者评论对图书的销量有显著的影响，也有人分析初次评论与追加评论对消费者购买决策的影响[22]。这类文献主要探讨评论的内容是否对商品的销量有影响及影响的不同侧面。

c）评语的情感分析[24~29]。如李一军和叶强等人[24,25]分析了商品在线评论情感倾向与商品销售收入的关系，如果获取的评论内容不符合分析的要求，那么情感倾向性的分析结果也将会有很大的误差[27]，王洪伟等人[28]在考虑人们表达习惯和语料粒度的基础上，提出一种基于句子情感的段落情感极性分类方法。王伟等人[29]认为在同类商品推荐中粗粒度识别评论情感极性，细粒度识别每个商品特征的情感极性。这类文献主要集中在主观性内容识别、褒贬情感分类以及在线评论的经济价值挖掘等几个方面。

已有的研究主要集中在消费者受评价时效性、激励措施、多次评论、语言特征等外界环境影响，或产品销量的影响，以及评语的正、负面倾向等内容进行讨论分析，鲜有消费者对商品的关注热点、评论语之间的关联和共现特征等方面的研究。本文在总结现有网购评语挖掘的基础上，试图通过复杂网络理论，找出网购消费者的关注热点，即通过网购消费者的评语，找出其中的高频词及高频词之间的共现网络，并对该网络的结构特性进行探讨。

1　中文网购评语高频词共现网络

本文所使用的评语数据取自全球前十大互联网公司之一的中国大型综合型电商平台，该平台 2014-2015 年累计销售额近 6 亿元。销售产品种类近 3150 万种，网购评语数据也非常庞大。本文仅选取该网站销售的 543 件衣服在 2014-2015 年共计 59730 条的评论数据进行分析。依据该 59730 条评论数据，构建网购评语的高频词共现网络。

1.1　中文分词和高频词

分词是数据预处理的关键步骤，是得到评语信息并剔除相关“的”“得”及各类标点符号后的第一个过程。常用的关于分词的方法有三种，即依赖于字符串匹配的分词方法、基于统计的分词方法、基于知识理解的分词方法。

考虑到数据较大，本文采用基于统计的分词方法，主要是在上下文中，分析相邻的字共同出现的次数，共同出现的次数越多构成词的可能就越大。也就是说字与字相邻出现的频率来决定是否构成词，它能较好地反映成词的可信度[30-31]。同时也可以利用现有中文分析软件分词和词频统计，如中科院开发的中文词法分析系统 ICTCLAS、武汉大学沈阳教授研发的 ROST CM 等。

分词后得到评论的词 c_i 和对应的词频数 n_i，该词在所有的 N 个评论中出现的概率 $p_i=n_i/N$，并将这些词按照词频数（或概率）从大到小排列，从而保证 $\forall i<j$ 时 $n_i \geq n_j$（等价于 $p_i \geq p_j$）。设置选取的高频词数 K，选取前 K 个词为高频词。

1.2　高频词共现矩阵和共现网络

K 个衣服评论语高频词中，统计任意两个高频词 c_i、c_j 在所有 N 个评论语中共现的次数为 e_{ij}。任意两个高频词之间的共现关系借用信息论中的互信息[32]来表示，描述两个词之间的关联程度。计算公式如下：

$$I_{i,j} = \log_2 \frac{P_{i,j}}{P_i P_j} \tag{1}$$

其中：$P_{i,j}$ 代表 c_i 和 c_j 同时出现的概率，P_i 表示 c_i 出现的概率，P_j 表示 c_j 出现的概率。分析可得 $I_{i,j}$ 结果越大，c_i 和 c_j 的共现关联程度也越大。由 $(I_{i,j})_{K\times K}$ 构成的矩阵即为高频词共现矩阵（考虑到对称关系 $I_{i,j}=I_{j,i}$，也可表示为高频词共现上三角或下三角矩阵）。

之所以选择互信息而不是选择高频词共现次数，主要原因可以用一个例子来解释。假设评论语有 10000 条，高频词 c_1、c_2 的词频数为 n_1=8000（即 P_1=0.8），n_2=7000(即 P_2=0.7)，c_1、c_2 共现的次数为 5000（即 $P_{1,2}$=0.5），互信息为 $I_{1,2}$=-0.36；高频词 c_3、c_4 的词频数为 n_3=5000（即 P_1=0.5），n_4=5000(即 P_2=0.5)，c_3、c_4 共现的次数为 4500（即 $P_{3,4}$=0.45），互信息为 $I_{3,4}$=0.85。虽然 c_3、c_4 的共现次数较小，但是相对各高频词而言，c_3、c_4 几乎都是同时出现，因此，c_3、c_4 更应该选作共现关系。在表示高频词之间的共现关系时，也可以借用关联规则中的置信度来取代本文中的互信息。

针对共现矩阵，按照互信息从大到小排列，设置高频词共现关系数 E，选取前 E 个高频词共现关系为边，这 E 条边涉及到的高频词为节点，形成的网络即为高频词共现网络。

例如，设置高频词数 K=200，构建高频词共现网络时选取共现关系数 E=100，得到的高频词共现网络如图 1 所示，节点的大小表示度的大小。

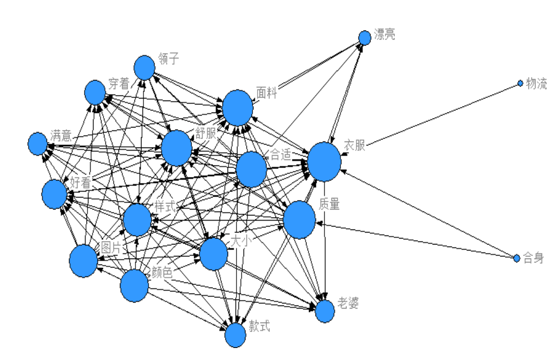


图 1　K=200，E=100 的高频词共现网络

由图 1 可知，共现网络的边数即为设置的 E 的数目，而这些边涉及的节点分别是来自 200 个高频词，即只有互信息排名前 100 的节点对，才会成为高频词共现网络的节点。

2　网购评语高频词共现网络的结构特性

2.1　中文分词和高频词

网购评语高频词共现网络的拓扑结构如图 1 所示。在网络中，一个节点代表一个网购衣服评语高频词，一条边表示两个网购衣服评语高频词之间同时出现在一个评语的共现关系。节点的大小表示该节点的度（degree），即连接到该节点的邻居节点的数量。一个节点的度是衡量其在网络中重要程度的关键性指标之一。有着高连接度的节点常常被称为高连通度节点或 hub 节点。不失一般性，假定网购衣服评语高频词共现网络中的节点数为 N，根据高频词共现网络的共现关系构建一个二元邻接矩阵 A(N,N)。如果高频词 i 到高频词 j 存在共现关系，矩阵 A(N,N)的元素 $a_{i,j}$ 取值为 1，否则取值为 0。对称矩阵 A(N,N)用来计算结构特性，如最短路径程度、网络密度、度分布、群聚系数、社区结构、富人俱乐部、匹配形式等。

2.2　网购评语高频词共现网络的小世界特性

许多现实世界的网络表现出小世界网络的结构特性，即与相同规模的随机网络相比，有着相同水平的平均路径长度和更高水平的聚集系数[33]。网络的平均路径长度是网络中所有节点对之间最短路径长度的均值，即

$$L=\frac{\sum_{i\neq j}d_{ij}}{N(N-1)} \tag{2}$$

其中: d_{ij} 是高频词节点 i 到高频词节点 j 所经历的边的数目。网络的聚集系数是网络中所有节点聚集系数的均值，即

$$C=\frac{1}{N}\sum_{i}\frac{N_i}{k_i(k_i-1)/2} \tag{3}$$

其中:k_i 是节点 i 的度，N_i 是 k_i 个邻居之间实际存在的边数。

根据以上定义，图 1 中的网购评语高频词共现网络中节点数 N 为 18，计算得该网络的平均路径长度为 1.34，聚集系数为 0.84。与相对应的随机网络相比，网购衣服评语高频词共现网络有着相同水平的平均路径长度和更高水平的群聚系数，表现出明显的小世界现象。此分析结果表明，网购衣服评语高频词共现网络中任何两个网购衣服评语高频词之间的共现关系至多平均需要一次过度，一半以上的高频词之间有着直接的共现关系，网购高频词与高频词之间有明显的共现关系。

根据网购评语高频词共现网络的小世界特性，可知任何两个高频词有着直接或间接的共现关系。结合图 1 的例子就可以为消费者的购买决策提供一定的建议，如消费者选择“舒服”这个高频词进行网上购物的检索，平台应自动搜索与“舒服”共现的高频词“好看”“合身”等进行个性化推荐。

2.3　网购评语高频词共现网络的度分布特性

网络节点的度分布特性是描述复杂网络结构特性的重要指标之一。现有文献中常用节点度的分布函数 P(k)或累积度分布函数 P(≥k)描述节点的度分布特性。前者的含义为复杂网络中度为 k 的节点数占总节点数的比例，后者的含义为复杂网络中度大于等于 k 的节点数占总节点数的比例[34]。实证研究表明，大量现实世界的复杂网络表现为三种类型的节点度分布特性：①无标度特性；②宽标度特性；③单标度特性。本文采用累积度分布函数描述网购衣服评语高频词共现网络的度分布特性，图 1 的网购衣服评语高频词共现网络的累积度分布函数如图 2 所示。

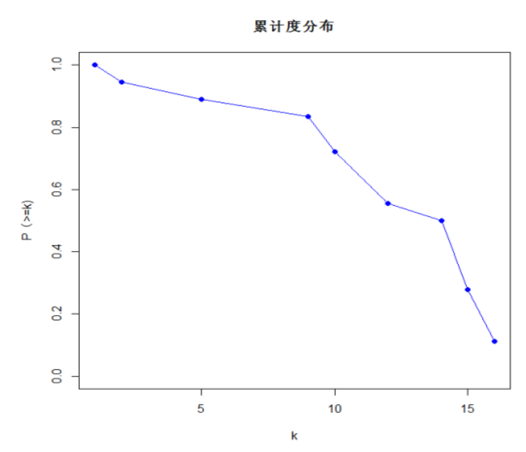


图 2　K=200，E=100 网购评语高频词共现网络的累积度分布

由于图 1 为 K=200 且 E=100 的高频词共现网络，避免网络不具有代表性，同时计算 K=500，E=200 的高频词共现网络（图 3）、K=1000，E=500 的高频词共现网络（图 4）的累积度分布函数如图 5 所示。

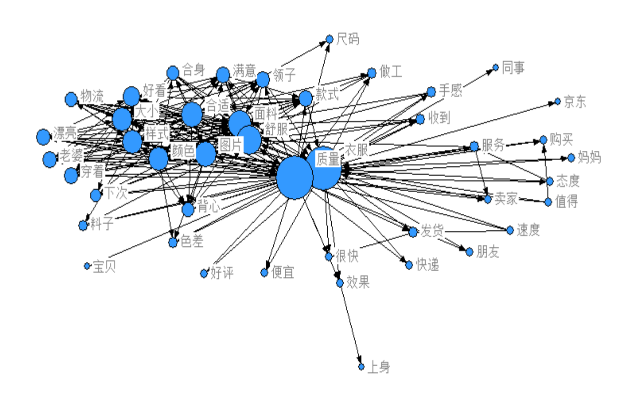


图 3　K=500，E=200 的高频词共现网络

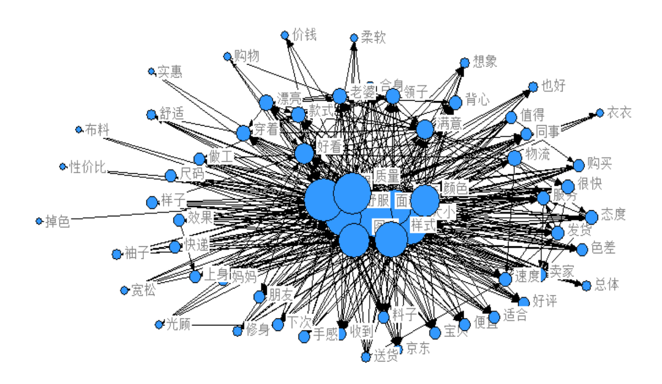
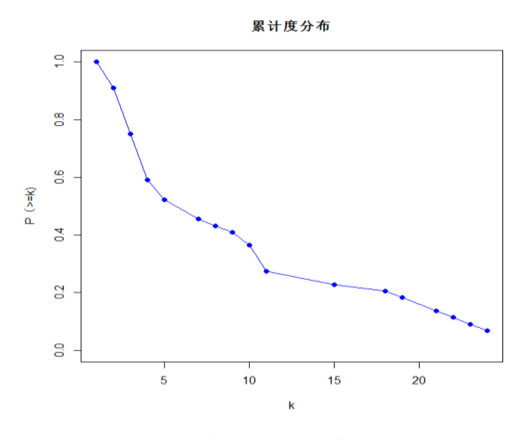
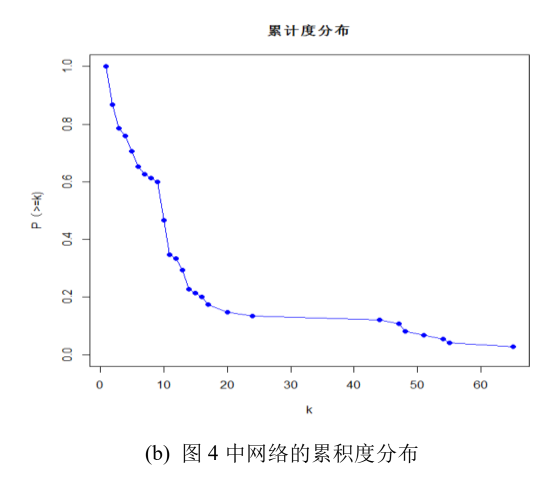


图 4　K=1000，E=500 的高频词共现网络



(a)图 3 中网络的累积度分布



(b) 图 4 中网络的累积度分布

图 5　累积度分布

由图 2、5 可以看出，网购衣服评语高频词共现网络的节点越多，随着度 k 的增大，累积度分布函数曲线表现出开始快速衰减、后期缓慢衰减的情形，说明该网络的节点度具有无标度特性。无标度特性表现为网络中少数节点的连接度较大（拥有极其多的连接），对网络的运行起着主导的作用，而大多数节点连接度较小（只有很少量的连接）。

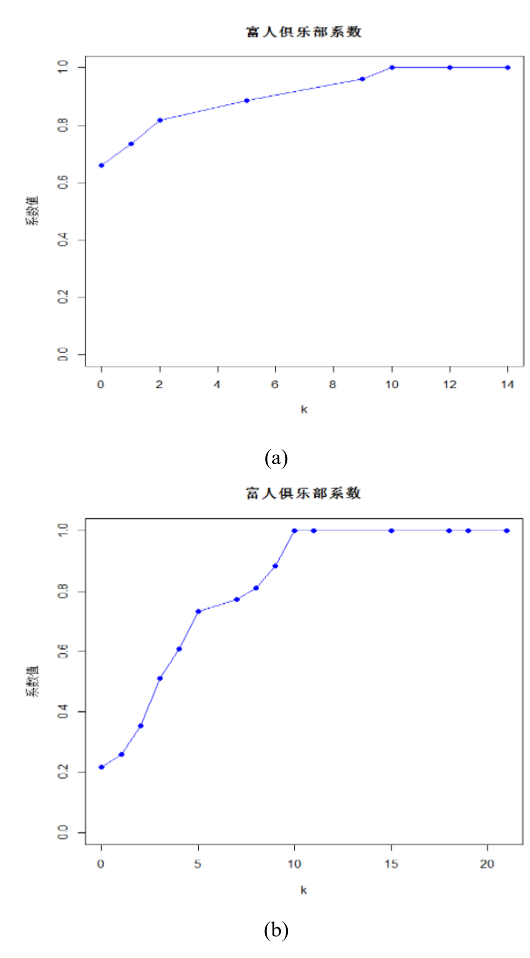
网购衣服评语高频词共现网络正好符合这一特点，网络中质量、颜色、样式、大小、面料等节点为关键节点，具有大量的连接，但是其他节点的连接较少。

2.4　网购评语高频词共现网络的富人俱乐部现象

富人俱乐部现象是指网络中连接度较大的节点（Hub 节点）之间有着紧密的连接，进而形成了网络中的一个核心团队，可以用富人俱乐部系数 $\varphi(k)$ 来度量[35]。用 $E_{>k}$ 表示网络中度大于 k 的节点之间的连接数量，富人俱乐部系数定义为

$$\varphi(k)=\frac{2E_{>k}}{N_{>k}\left(N_{>k}-1\right)} \tag{4}$$

其中: $N_{>k}\left(N_{>k}-1\right)/2$ 表示度大于 k 的节点之间最大可能的连接数量。网购评语高频词共现网络的富人俱乐部系数如图 6 所示，图 1、3 对应的富人俱乐部系数如图 6 中的（a）（b）。可以看出，该系数随着节点度 k 的增大而增大，意味着网购评语高频词共现网络中 hub 评语之间有着比其他低连接度节点更加紧密的连接，形成了一个富人俱乐部。同时可以看出，度大于 10 的网购评语节点形成了全连通图。网购评语高频词共现网络的富人俱乐部现象说明，处于俱乐部中的评语是该网络的核心，控制着整个网络上的评语节点构成。因此，网购消费者对这些 hub 评语节点的关注程度及高，这些方面对吸引消费者购买起着至关重要的作用。



(a)

(b)

图 6　富人俱乐部系数

2.5　匹配形式

匹配形式描述了网络的节点度与其邻居节点度之间的关系[34]。统计上分析网络的匹配形式包括两个步骤。首先，计算节点 i 的邻居平均度如式（5）所示。

$$k_{nn}\left(i\right)=\frac{1}{k_i}\sum_{j\in N_i}k_j \tag{5}$$

其中：N_i 为节点 i 的邻居集合。然后，对具有相同度 k 的节点的邻居平均度进行统计平均，即 $\bar{k}_{nn}\left(k\right)=\frac{1}{N_k}\sum k_{nn}\left(i\right)$，其中 N_k 是网络中度为 k 的节点的数量。如果 $\bar{k}_{nn}(k)$ 随着 k 的增大而增大，意味着高连通度的节点偏好与其他高连通度的节点相连，则网络表现为同配性。否则，如果 $\bar{k}_{nn}(k)$ 随着 k 的增大而降低，则网络表现为异配性。图 8 给出网购评语高频词共现网络的节点度与其邻居节点度之间的关系。可以看出，$\bar{k}_{nn}(k)$ 随着 k 的增大而减小，说明网购评语高频词共现网络是异配网络。此分析结果表明，在网购评语高频词共现网络中，高连通度的节点偏好与低连通度的节点相连。这进而说明在其生成演化过程中，

新进入网络中的评语优先与那些已经有着较高连接度的个体建立连接，现有评语之间共现关系的建立也偏好于发生在高连通度评语和低连通度评语之间。

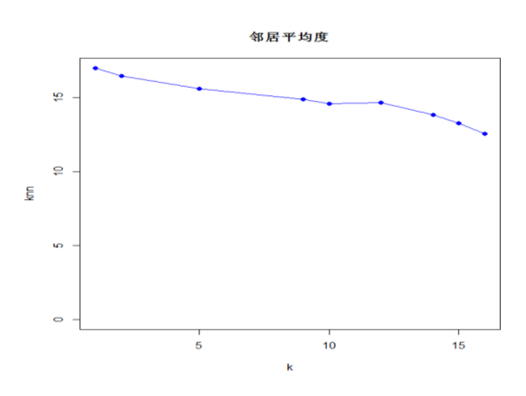


图 8　相同度 k 的节点的邻居平均度 $\overline{k}_{nn}(k)$

2.6　网购评语高频词共现网络的社区结构

社区结构是指网络中存在若干群，群内的点连接紧密，群之间的点联系稀疏。本文使用 Girvan 和 Newman 定义的模块性函数寻找网购衣服评语高频词共现网络中的社区结构[36]。借鉴 GN 算法中的模块性指标 Q，将网络划分为 c 个子网，再定义一个对称矩阵 $E=(e_{ij})_{c\times c}$,其中 e_{ij} 表示网络中连接两个子网 i 与 j 的节点的边占整个网络中的边的比例。矩阵中对角线各元素之和 $t=\sum_{i=1}^{c}e_{ii}$,它描述网络之中连接子网内部节点之间的边占整个网络的边的比例。$b_i=\sum_{j=1}^{c}e_{ij}$,它描述与第 i 个子网中的节点有连接的边占整个网络之中边的比例。从而，得到模块性函数 Q。

$$Q=\sum_{i=1}^{C}\left(e_{ii}-b_i^2\right)=t-\|E^2\| \tag{6}$$

其中：$\|E^2\|$ 表示矩阵 E^2 中的各元素之和。

如果子网划分效果越佳，Q 值一般较大。但是，Q 永远不能超过 1，Q 越靠近 1，则表明子网的结构越容易明显地辨别。而通常在实际网络中，Q 的实际值常常介于 0.3 与 0.7 之间。根据该定义，可以得知 Q 值是可以为负的。当子网结构内部边所占比例小于任意连接时边所占比例期望值才发生这种情况。在划分子网络的过程当中，通过计算各种划分情况的模块性 Q 值，找到 Q 值最大时所对应的相应子网划分，那么这样的子网划分就是最佳或者最接近最佳的划分方式。

本文网购衣服评语高频词共现网络的 $Q=-0.082$ ，网络被分为 4 个子网。该网络具有的最大模块性指标值为负值，远远低于 0.3，说明它所给出的 4 个子网的划分实际上是不明显的，其社区结构不显著。

3　结束语

基于复杂网络理论，将网购评语的关联抽象成网络形式，即以网购评语高频词为节点，高频词共现关系为边，建立了网购衣服评语高频词共现网络，之后对其结构特性进行研究，得到了以下主要结论：

a）网购评语高频词共现网络是一个小世界网络，任何两个评语之间的共现关系至多平均需要一次中转，一半以上的评语高频词之间有着直接的共现关系，形成一个高频词共现网络。

b）网购评语高频词共现网络的度分布形式为无标度分布，网络中少数节点的连接度较大，对网络的运行起着主导的作用，而大多数节点连接度较小，说明网购用户集中关注的因素较为集中。

c）网购评语高频词共现网络具有富人俱乐部现象，处于俱乐部中的高频词是网购评语高频词共现网络中的核心词，是整个评语网络上的高专注评语，它们决定了网购消费者对网购消费的关注因素。

d）网购评语高频词共现网络具有节点度的异配特性，高连通度的节点偏好与低连通度的节点相连。

e）网购评语高频词共现网络不存在明显的社区结构，其模块性指标 Q 值为负（-0.082），这意味着，网购评语高频词共现网络中的所有高频词形成了紧密的连接。

此外，本文的研究方法为电子商务零售企业全面地了解自身的产品和用户对产品的期待提供了方法支持，也为消费者制定购买策略、管理部门制定监管策略提供了支持。接下来，基于网购评语高频词共现网络的结构特性对电商零售企业制定精准营销的策略进行分析，探讨网购评语高频词共现网络结构特性与消费者网购行为、商品销量、商品销售热度之间的关系，将是非常有意义的工作。

参考文献：

[1] 杨铭，祁巍，闫相斌，等. 在线商品评论的效用分析研究 [J]. 管理科学学报, 2012, 15 (5): 65-75.

[2] 廖成林，蔡春江，李忆. 电子商务中在线评论有用性影响因素实证研究 [J]. 软科学, 2013, 27 (5): 46-50.

[3] 聂卉. 基于内容分析的用户评论质量的评价与预测 [J]. 图书情报工作, 2014, 58 (13): 83-89.

[4] 姜巍，张莉，戴翼，等. 面向用户需求获取的在线评论有用性分析 [J]. 计算机学报, 2013, 36 (1): 119-131.

[5] Titov I, Mcdonald R. Modeling online reviews with multigrain topic models [C]// Proc of the 17th International Conference on World Wide Web. 2008: 111-120.

[6] Yang L, Xiangji H, Aijun A, et al. Modeling and predicting the helpfulness of online reviews [C]// Proc of the 8th IEEE International Conference on Data Mining. 2008: 443-452.

[7] Cheung C M, Lee M K. What drives consumers to spread electronic word of mouth in online consumer-opinion platforms [J]. Decision Support Systems, 2012, 53 (1): 218-225.

[8] 闫强，孟跃. 在线评论的感知有用性影响因素——基于在线影评的实证研究 [J]. 中国管理科学, 2013, 21: 126-131.

[9] 付东普，王刊良．评论回报对在线产品评论的影响研究——社会关系视角 [J]. 管理科学学报, 2015, 18 (11): 1-12.

[10] 孟美任，丁晟春．虚假商品评论信息发布者行为动机分析 [J]. 情报科学. 2013, 31 (10): 100-104.

[11] 王长征，何钐，王魁．网络口碑中追加评论的有用性感知研究 [J]. 管理科学. 2015, 28 (3): 102-114.

[12] 殷国鹏．消费者认为怎样的在线评论更有用?——社会性因素的影响效应 [J]. 管理世界. 2012, 12: 115-124.

[13] 王君珺，闫强．不同热度搜索型产品的在线评论对销量影响的实证研究 [J]. 中国管理科学, 2013, 21: 406-411.

[14] 李宏，喻葵，夏景波．负面在线评论对消费者网络购买决策的影响：一个实验研究 [J]. 情报杂志, 2011, 30 (5): 202-207.

[15] 冯娇，姚忠．基于社会学习理论的在线评论信息对购买决策的影响研究 [J]. 中国管理科学, 2016, 24 (9): 106-114.

[16] Li Xinxin, Hitt L M. Self-selection and information role of online product reviews [J]. Information Systems Research, 2008, 19 (4): 456-474.

[17] Mudambi S M, Schuff D. What makes a helpful online review?A study of customer reviews on amazon. com [J]. MIS Quarterly, 2010, 34 (1): 185-200.

[18] Kwark Y, Chen Jianqing, Raghunathan S. Online product reviews: implications for retailers and competing manufacturers [J]. Information Systems Research, 2014, 25 (1): 93-110.

[19] 章璇，景奉杰．网购商品的类型对在线冲动性购买行为的影响 [J]. 管理科学, 2012, 25 (3): 69-77.

[20] 张黎．网民传统的网下购物习惯与其网上购物行为的关联性及其实践意义 [J]. 管理科学. 2006, 19 (2): 13-21.

[21] 龚诗阳，刘霞，赵平．线上消费者评论如何影响产品销量?——基于在线图书评论的实证研究 [J]. 中国软科学, 2013, 6: 171-173.

[22] 石文华，龚雪，张绮，等．在线初次评论与在线追加评论的比较研究 [J]. 管理科学, 2016, 29 (4): 45-58.

[23] 杜学美，丁璟妤，谢志鸿，等．在线评论对消费者购买意愿的影响研究 [J]. 管理评论, 2016, 28 (3): 173-183.

[24] 张紫琼，叶强，李一军．互联网商品评论情感分析研究综述 [J]. 管理科学学报, 2010, 13 (6): 84-96.

[25] 郝媛媛，邹鹏，李一军，等．基于电影面板数据的在线评论情感倾向对销售收入影响的实证研究 [J]. 管理评论, 2009, 21 (20): 95-103.

[26] 李永忠，胡思琪．基于 HowNet 和 PAT 树的网购评语情感分析 [J]. 图书情报研究, 2016, 3: 66-70.

[27] Hu M, Liu B. Mining and summarizing customer reviews [C]// Proc of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2004: 168-171.

[28] 王洪伟，郑丽娟，尹裴，等．基于句子级情感的中文网络评论的情感极性分类 [J]. 管理科学学报, 2013, 16 (9): 64-74.

[29] 王伟，王洪伟，孟园．协同过滤推荐算法研究：考虑在线评论情感倾向 [J]. 系统工程理论与实践, 2014, 34 (12): 3238-3249.

[30] 蒋建洪，赵嵩正，罗玫．词典与统计方法结合的中文分词模型研究及应用 [J]. 计算机工程与设计, 2012, 33 (1): 387-391.

[31] 张敏，王春红．基于统计力法的 web 新词分词方法研究 [J]. 计算机工程与科学, 2010, 32 (5): 133-135.

[32] 黄德根，马玉霞，杨元生．基于互信息的中文姓名识别方法 [J]. 大连理工大学学报, 2004, 44 (5): 744-748.

[33] Watts D J, Strogatz S H. Collective dynamics of‘small-world’networks [J]. Nature, 1998, 393 (4): 440-442.

[34] Newman M. E. J. The structure and function of complex networks [J]. SIAM Review, 2003, 45 (2): 167-256.

[35] Zhou S, Mondragon R J. The rich-club phenomenon in the Internet topology [J]. IEEE Communications Letters, 2004, 8 (3): 180-182.

[36] Girvan M, Newman M E J. Community structure in social and biological networks [C]// Proc of National Academy of Sciences of the USA, 2002, 99 (12): 7821-7826.